FEBS Open Bio 5 (2015) 325–334

FEBS openbio

journal homepage: www.elsevier.com/locate/febsopenbio

Novel insights into structure–function mechanism and tissue-specific expression profiling of full-length *dxr* gene from *Cymbopogon winterianus*

Kamalakshi Devi [a], Budheswar Dehury [a,b], Munmi Phukon [a,c], Mahendra Kumar Modi [a], Priyabrata Sen [a,*]

[a] Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat 785013, Assam, India
[b] Regional Medical Research Centre (ICMR), Chandrasekharpur, Nandankanan Road, Bhubaneswar 751023, Odisha, India
[c] School of Biological Sciences, College of Life Science & Medicine, University of Aberdeen, Aberdeen AB24 3UU, Scotland, UK

ARTICLE INFO

Article history:
Received 2 January 2015
Revised 9 April 2015
Accepted 10 April 2015

Keywords:
Citronella
DXR
Molecular modelling
Molecular dynamics simulation
Docking

ABSTRACT

The 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR; EC1.1.1.267), an NADPH-dependent reductase, plays a pivotal role in the methylerythritol 4-phosphate pathway (MEP), in the conversion of 1-deoxy-D-xylulose-5-phosphate (DXP) into MEP. The sheath and leaf of citronella (*Cymbopogon winterianus*) accumulates large amount of terpenes and sesquiterpenes with proven medicinal value and economic uses. Thus, sequencing of full length *dxr* gene and its characterization seems to be a valuable resource in metabolic engineering to alter the flux of isoprenoid active ingredients in plants. In this study, full length DXR from citronella was characterized through *in silico* and tissue-specific expression studies to explain its structure–function mechanism, mode of cofactor recognition and differential expression. The modelled DXR has a three-domain architecture and its active site comprised of a cofactor (NADPH) binding pocket and the substrate-binding pocket. Molecular dynamics simulation studies indicated that DXR model retained most of its secondary structure during 10 ns simulation in aqueous solution. The modelled DXR superimposes well with its closest structural homolog but subtle variations in the charge distribution over the cofactor recognition site were noticed. Molecular docking study revealed critical residues aiding tight anchoring NADPH within the active pocket of DXR. Tissue-specific differential expression analysis using semi-quantitative RT-PCR and qRT-PCR in various tissues of citronella plant revealed distinct differential expression of DXR. To our knowledge, this is the first ever report on DXR from the important medicinal plant citronella and further characterization of this gene will open up better avenues for metabolic engineering of secondary metabolite pathway genes from medicinal plants in the near future.

1. Introduction

Citronella (*Cymbopogon winterianus*), an aromatic grass with large reserves of monoterpenes and sesquiterpenes, is grown for commercial and industrial purposes in tropical as well as subtropical regions of Asia, America and Africa [1]. Industrial interest in essential oils is due to their application as fragrance in perfumes, as flavor additives for use in food products or even as pharmaceutical products [2]. Citronella is traditionally known for its insect-repellent nature, especially against malaria-causing mosquito species, therapeutic properties against various diseases [3–5] and its anti-fungal property [6]. Strong antifungal activity against several species of *Aspergillus*, *Penicillium* and *Eurotium* due to presence of citronellal and linalool, components of citronella oil, has also been reported [7]. Citronella oil is used to expel worms or other parasites from the intestines, control muscle spasms, increase appetite, and increase urine production (as a diuretic) to relieve fluid retention [8].

The Java citronella (*C. winterianus*) is known for its high quality and quantity of essential oils ranging from 0.8% to 1.0% of total weight having constituents like limonene (1.8%), geraniol (~23%), elemol (~10%), geranyl acetate (4.0%), α-cadinol (8.0%), citronellol (~10%), citronellal (~35%) etc., [9]. The processors for all these isoprenoid compounds are two common 5-carbon molecules i.e., isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In higher plants, these two isoprenoid compounds are synthesized by two non-related biosynthetic pathways, from different precursors [10–13]. In the mevalonate (MVA) pathway, IPP and DMAPP are synthesized from mevalonic acid whereas,

* Corresponding author. Tel.: +91 376 2340001 (O); fax: +91 376 2340101.
E-mail address: pbsen14@yahoo.co.in (P. Sen).

http://dx.doi.org/10.1016/j.fob.2015.04.005
2211-5463/

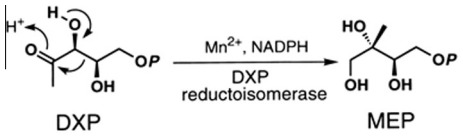

Fig. 1. 1-Deoxy-D-xylulose-5-phosphate reductoisomerase reaction (Adopted from Takahashi et al. (1998) [23]).

pyruvate and D-glyceraldehyde 3-phosphate are used as precursors for IPP and DMAPP synthesis via mevalonate-independent plastidial methylerythritol 4-phosphate (MEP) pathway [14–20].

The MEP pathway is composed of seven enzymatic steps. But the second step i.e., NADPH-dependent reduction and conversion of 1-deoxy-D-xylulose-5-phosphate (DXP) into MEP by intramolecular rearrangement is considered as most crucial [21–23] (Fig. 1). This reaction is catalyzed by 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR: EC1.1.1.267); which is the first committed step of MEP pathway [24]. It has been seen that disruption of the 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) gene in Arabidopsis results in albino, dwarf and defects in trichome initiation and stomata closure due to insufficient supply of all the major plant hormones synthesized from IPP and DMAPP [25]. Their results reveal a critical role for the MEP biosynthetic pathway in controlling the biosynthesis of isoprenoids. *DXR* gene of *Zymomonas mobilis* has been characterized which shows highest percentage of sequence identity (i.e., 48.2%) with *dxr* of *Escherichia coli* [26]. Recently, several groups concentrated on molecular cloning and characterization of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene from various plant species *viz*., *Ginkgo biloba* [27], *Elaeis guineensis* Jacq. [28], *Camptotheca acuminate* [29] and *Salvia miltiorrhiza* [30]. The expression and molecular analysis of the *Arabidopsis dxr* gene along with the comparative analysis of all plant DXR sequences known to date, where it possess an N-terminal transit peptide for plastids, with a conserved cleavage site, and a conserved proline-rich region at the N terminus of the mature protein, which is absent in prokaryotic DXR homologs. It has also been reported that *Arabidopsis* DXR participates in the control of 2-C-methyl-D-erythritol 4-phosphate pathway [31].

The DXR is a key enzyme in MEP pathway and therefore a change in DXR activity could alter the flux of isoprenoid expressed down the pathway. Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase improves flux through the MEP pathway that leads to increased monoterpene production [32,33]. The sheath and leaf of *C. winterianus* also accumulates large amount of terpenes and sesquiterpenes, compound with proven medicinal value and economic uses, so sequencing of full length *dxr* gene and its characterization would enable prospective manipulations of the pathway that may lead to increased accumulation of active ingredients. However, the unavailability of X-ray- or nuclear magnetic resonance (NMR)-derived structures of DXR has hindered the understanding of three-dimensional architecture, mode of cofactor recognition and critical residues participate in catalysis. Therefore in the present study, we have utilized the full length *dxr* gene sequence (GenBank Accession number: KJ749651) from one of our current citronella leaf whole transcriptome sequencing study (BioProject ID: PRJNA263976). In this study we have used a combinatorial approach involving homology modelling, molecular dynamics simulations and docking to explain the 3-dimensional domain architecture, dynamics behaviour and mode of cofactor recognition in DXR to facilitate the understanding of catalysis of *dxr* in citronella. Further, tissue specific differential expression analysis of DXR was performed both by semi-quantitative RT-PCR and quantitative real time PCR in various tissues of citronella plant. The results from the present study are expected to open up better avenues for metabolic engineering of important secondary metabolite pathway genes from important medicinal plants like citronella, and also in other related medicinal plant species.

2. Results and discussion

2.1. Sequence analysis

The domain prediction of DXR by SMART against Pfam database revealed two putative domains *viz*., 1-deoxy-D-xylulose 5-phosphate reductoisomerase (Pfam: DXP_reductoisom) positioned between 81 and 209 amino acid and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (Pfam: DXP_redisom_C) comprised of 223–306 amino acids. The mature 465 amino acid peptide is neutral in nature (pI of 6.44) with a molecular mass of 50.8 kDa. The aliphatic index of DXR is very high (96.5) indicating the stability of the protein over a wide range of temperature. Instability index was 35.1 (<40) DXR is classified as a stable protein. The GRAVY index of DXR is very low (0.056) indicating the possibility of its better interaction with water. To characterize the sequence of the N-terminal region of CwDXR, we aligned CwDXR with the same region of the plant DXRs known to date (Fig. 2). The plant enzyme contains an extension of 73–80 residues that is not present in the prokaryotic sequence [24]. The prediction of transit peptide for plastids with the ChloroP Program [34] in CwDXR sequence envisaged transit peptide processing site at the N terminus of a conserved Cys-Ser-X motif (49–51 position), where X means any of the hydrophobic residues Ala, Val, or Met (Met51) in CwDXR). At the N-terminal end, the sequence is weakly conserved but enriched in Ser residues, features that are typical of plastid transit peptides [35]. In contrast, the extended region at the N-terminal side (positions 52–80 of CwDXR) is highly conserved and specifically rich in Proline residues (Fig. 2). The number of Pro residues in this region ranges from 6 to 8. The consensus motif P(P/Q) PAWPG(R/T) A can be defined in the Proline rich region of plant DXR (positions 55–62 of CwDXR sequence). The primary sequence analysis suggests that CwDXR has a transit peptide for plastids, is processed at a conserved cleavage site, and contain an extended Pro-rich region at N-terminus of the mature protein. This consensus P(P/Q)PAWPG(R/T) Pro-rich motif of all plant DXR sequences may be used as a signature for DXR protein. Analysis of secondary structure elements through CONCORD revealed random coils (55.47%) dominates over helices (33.33%) followed by strands (11.20%).

2.2. Phylogenetic analysis

To characterize the full length sequence of DXR in citronella, the polypeptide was aligned with few other homologs' plant DXRs known to date. Multiple sequence alignment of DXR with its closest homologs revealed that initial region of the N-terminal of all the selected sequences(1–55 amino acids) are not conserved whereas, the central region and C-terminal region is strongly conserved. The 2-D phylogenetic tree inferred from NJ method revealed dichotomy with two distinct clusters (Fig. 3). Cluster-I comprised of DXR from monocots including citronella whereas, dicots formed the second cluster (Cluster-II) well supported by strong bootstrap values within their nodes. So, it can be inferred that DXR gene may have evolved due to parallel duplication from the same ancestral gene which has transferred subsequently to different lineages of plant species. The closest homology of citronella DXR has been found with *Zea mays* (97% identity) followed by *Seteria italica* (96% identity) which formed distinct cluster (Cluster I).

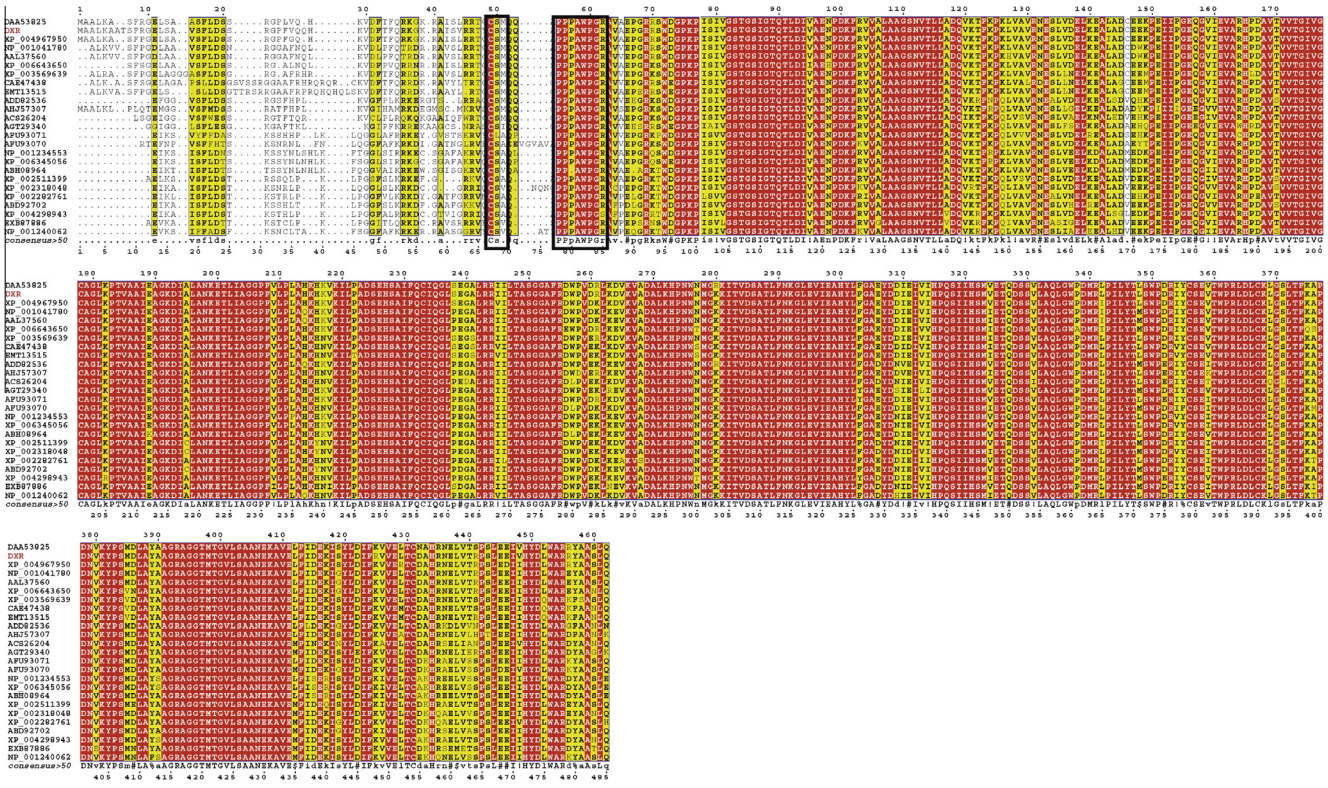

Fig. 2. Multiple sequence alignment of CwDXR sequence with all the available plant DXR sequences using MEGA v6.1 and ESPript. Conserved regions are highlighted in red square boxes labelled in white whereas partially or conserved residues are marked in yellow square boxes labelled in black.

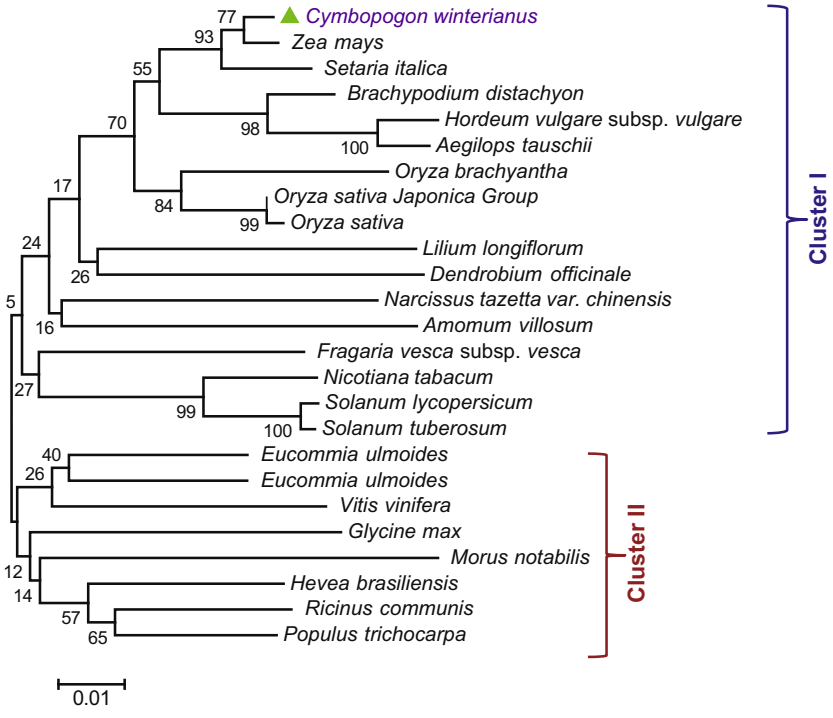

Fig. 3. Phylogenetic analysis of DXR from citronella with its closest homologs using Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the *p*-distance method and are in the units of the number of amino acid differences per site. Evolutionary analyses were conducted in MEGA6.

2.3. Structure–function studies

There were three top hits of DELTA-BLAST search i.e., 1R0K-A, 1Q0L-A and 3IIE-A which were considered as putative templates for DXR (Table 1). Out of the three templates, 1R0K-A (DXR of *Z. mobilis* [36]) showed greater sequence identity 45% with coverage 83% over other templates. Pairwise alignment of target with closest homolog (1R0K) using Multalign Program portrayed the N-terminal (1-74 residues) didn't show any significant homology with target, so this region was not taken into consideration for further model building exercises (Fig. 4).

2.4. Comparative modelling and model validation

The optimised model after a round of energy minimization was assessed for residue-by-residue stereo-chemical quality evaluation using various model validation servers. Ramachandran plot analysis using Procheck revealed that the DXR model had 93.3% Φ and Ψ angles of its residues within the core region of the Ramachandran plot as compared with 93.5% of the template, which signifies the accuracy of modelled DXR (Fig. 5). All the bond distances and angles of modelled protein lie within the allowed range of the standard dictionary values indicating that DXR model is reasonably good in geometry and stereochemistry. The Φ and Ψ distribution of Ramachandran plot of non-glycine, non-proline residues in the model and template (1R0K-A chain) are summarized in Table 2. The packing quality of each residue of the model was assessed by Verify3D Program, where the compatibility of the model residues with their environment is assessed by a score function. Residues with a score over 0.2 should be considered reliable. The score of the refined model maximally was above 0.2 which corresponds to acceptable side chain environment (Table 2). ProSA revealed a Z-score of −9.62 for modelled (Z-score of ZmDXR is −9.68) reflecting the overall quality of the DXR model. To check the degree of structural similarity of modelled target and template, we have measured the RMSD between equivalent Cα atom pairs along with a pair-wise 3-D alignment search of the template protein with the modelled structure through iPBA and MATRAS, which showed an overall identity of mere 94.70% for 375 aligned residues with a very low RMSD of 1.08 Å. The results of iPBA and MATRAS web server are in agreement with DXR and its structural homolog sharing strong structural conservation and similar in terms of structural

Table 1
Templates selected for combative modelling of DXR from citronella.

PDB ID	Source	Total score	Query coverage (%)	Sequence identity (%)	*E*-value	Resolution (Å)
1R0K-A	*Zymomonas mobilis*	388	83	45	3e-107	1.91
1Q0L-A	*E. coli*	406	75	45	2e-100	2.61
3IIE-A	*Yersinia pestis*	401	83	45	2e-100	2.21

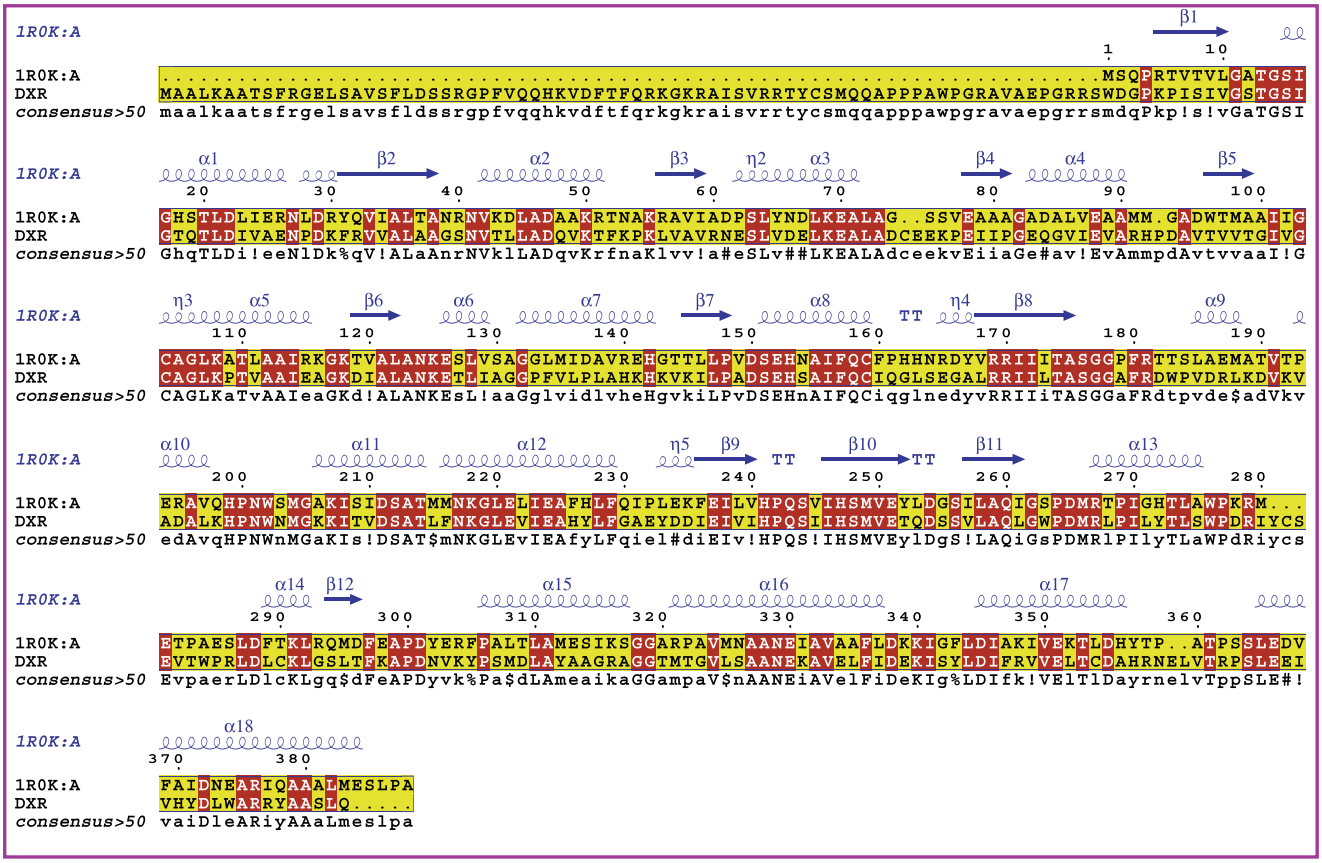

Fig. 4. The pair-wise sequence alignment of the target DXR and template 1R0K was constructed using Multalign and ESPript. The secondary structural elements were identified from the 1R0K structure using ESPript. The α-helices, η-helices, β-sheets and strict β-turns are denoted α, η, β and TT respectively. Similar amino acids are highlighted in boxes, and completely conserved residues are indicated by white lettering on a red background.

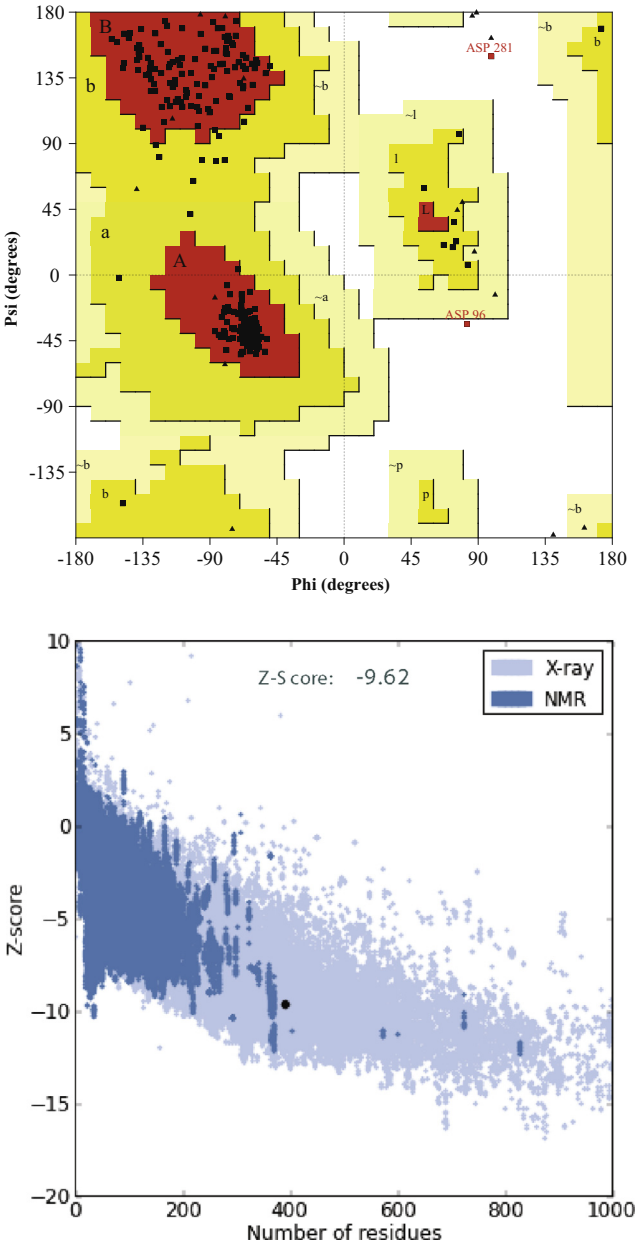

Fig. 5. Ramachandran plot and ProSA analysis of modelled DXR obtained using Procheck Program and ProSA Web tool.

folding. It also signifies that the generated model is reasonably good for further studies. From the above analysis, it is evident that the geometric quality of the backbone conformation, the residue contact, the residue interaction, and the energy profile of the structure are well within the limits which confirms the reliability of the modelled structure.

2.5. Molecular dynamics simulation of modelled DXR

The refinement of the modelled protein was subjected to MD simulation to get an optimized and stable structure suitable for docking with cofactor. The stability and dynamic properties of DXR model was observed by MD simulation. For cross-comparison and to understand the dynamics behaviour of the modelled DXR and its closest structural homolog, we employed MD simulation for 10 ns in explicit water for both the systems. To understand the stability of both the systems, the RMSD of Cα backbone atoms,

Table 2
Comparison of Ramachandran plot statistics of DXR model with closest structural homologue 1R0K.

Ramachandran plot statistics	Modelled structure of DXR		Template (1R0K-A)	
	Residues	Percentage	Residues	Percentage
Residues in most favoured regions	319	93.3	1262	93.5
Residues in additionally allowed regions	21	6.1	18	5.3
Residues in generously allowed regions	0	0.0	1	0.3
Residues in disallowed regions	2	0.6	1	0.3
Number of non-glycine and non-proline	342	100	338	100
Number of end residues (excluding Gly and Pro)	2	–	2	–
Number of glycine residues	26	–	22	–
Number of proline residues	21	–	18	–
Overall G factor	−0.03	–	0.42	–

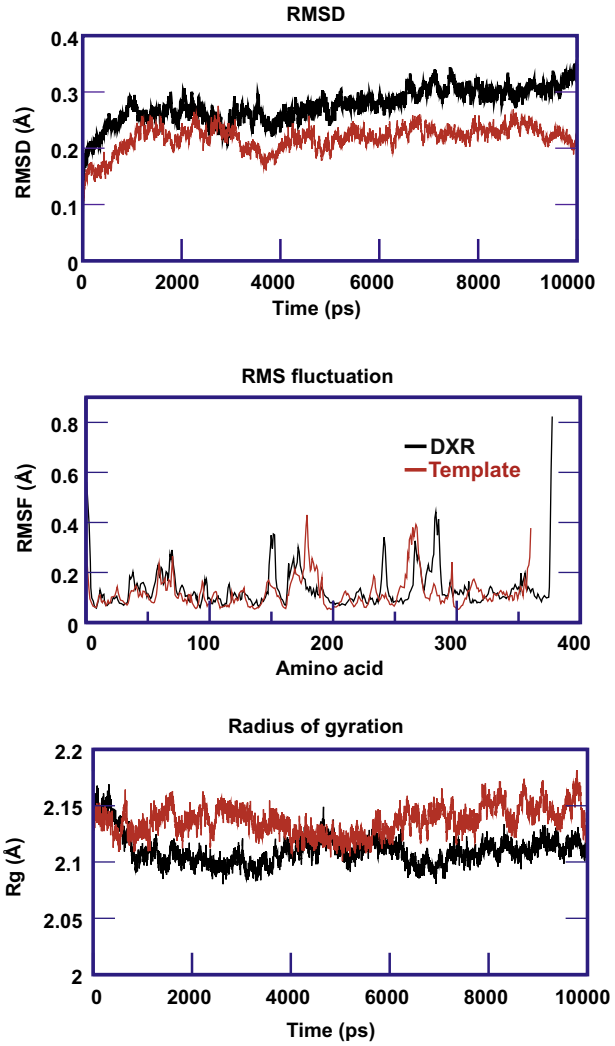

Fig. 6. The RMSD, RMSF and radius of gyration graph of the modelled CwDXR and template 1R0K: A (ZmDXR) during MD simulation. (a) RMSD of backbone Cα atoms of the CwDXR modelled structure and ZmDXR. (b) RMSF analysis of amino acid residues of CwDXR model structure and ZmDXR. (c) Radius of gyration of CwDXR modelled structure and ZmDXR during 10 ns trajectory. All the images were generated using XMGRACE software. Black and red colors represent CwDXR and ZmDXR models, respectively.

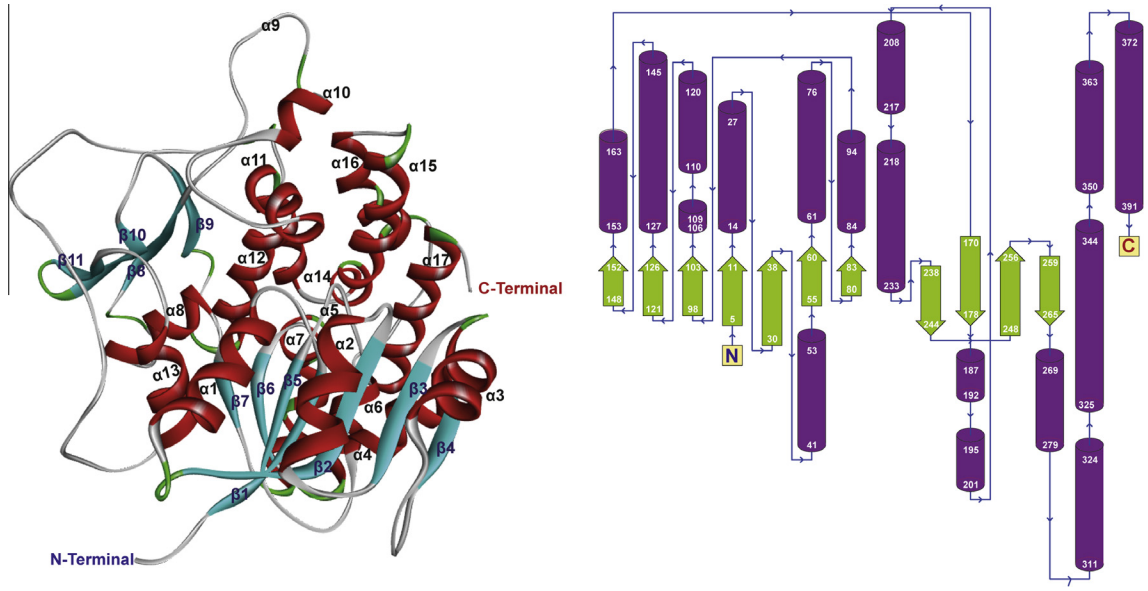

Fig. 7. Homology model of DXR from citronella and its Topology. Solid ribbon representation of the DXR model coloured by its secondary structure elements. The secondary structure elements such as α helices, β strands and the N and C termini are labelled. The images were prepared with Discovery Studio and Profunc Web server.

was calculated. The calculated RMSD was found to be ~0.3 Å; whereas, the template had an average RMSD of 0.24 Å. After initial 4 ns, the system remained in the plateau state till the 10 ns. Based on the intrinsic dynamics, and improved relaxation of the model protein, the potential energy and total energy of the structure was calculated and the radius of gyration graph was plotted with respect to different time scale. As compared to the template, DXR model had a less radius of gyri. The protein remained within compact radius of gyri of 2.2 Å indicating stability nature of the protein (Fig. 6). RMSFs of Cα atoms of the model and the template showed similar type fluctuations throughout the simulation process with minute deviation at the N-terminal end. As evidenced from Fig. 6, it can be clearly observed that the (amino acids positioned in 150–300 within the catalytic domain) residues showed more deviation reflecting flexibility of this region in protein. In addition, energy profile of the both systems i.e., the total and potential energy of model as compared the template indicated the dynamic stability of the proposed model. To assess changes in the structural arrangement of elements during 10 ns simulation, each residue was minutely observed and secondary structure elements at different time scale were mapped. As evident from Fig. 6, that most of the secondary structure elements of modelled DXR remained stable which highlight the stability and reliability of the DXR model for docking studies. A representative structure was extracted from stable of portion of trajectory (last 2 ns) to study the 3-D architecture of DXR protein.

2.6. Structural study of DXR protein

Detailed structural study of DXR protein was performed to understand the arrangement of secondary structural elements in 3-D space. Modelled DXR comprised of 17 α-helices (51.15% of amino acid, 200 residues), 11 beta-strands (16.11% of amino acid, 63 residues) and other secondary structure elements (turns/coils: 32.99%, 128 residues) (Fig. 7). So as to understand the active site architecture of modelled DXR, the pair-wise structural superposition of DXR with its closest structural homolog i.e., DXR of *Z. mobilis* was extensively studied. The active site of modelled DXR also consists of the two pockets i.e., NADPH binding pocket and the substrate binding pocket as also reported in ZmDXR [3]. The modelled DXR bears a three domain architecture *viz.*, N-terminal domain (residues 1–152) with 7 β-sheets and 7 α-helices, catalytic domain (residues 153–279) with 4 β-sheets and 6 α-helices and small C-terminal domain (residues 311–391) with only 4 α-helices. This domain distribution resembled the DXR from *Z. mobilis*, which is composed of three domains the N-terminal domain, residues 1–150, the catalytic domain comprising residues 151–299 and the C-terminal domain, residues 300–386 [32]. The N-terminal domain contains a seven-stranded parallel β-sheet, with three α-helices on each side of the sheet. The catalytic domain also belongs to the class of α/β folds; composed of a four-stranded β-sheet, and 6 α-helices and the smaller C-terminal domain consists of a bundle of four α-helices. Interestingly, the electrostatic surface potential analysis of both DXR from citronella and *Z. mobilis* revealed insightful variations in the charge distribution over the substrate binding site, which can be correlated with the sequence variability (variation in the key catalytic residues i.e., evident from the pair-wise structural superposition of the modelled DXR and crystal structure) and may suggest distinct substrate-binding patterns and differences in the catalytic mechanism.

2.7. Docking of co-factor NADPH

The residue conservation of the binding site and structural comparisons of NADPH-dependent DXR with known NADPH-dependent forms are crucial for predicting the cofactor specificity and the enzymatic mechanism. Molecular docking was performed to understand the recognition specificity and mode of cofactor binding to DXR using CDOCKER module of DS3.5. The consensus scoring system i.e., CDOCKER ENERGY, CDOCKER Interaction Energy, Ligscore1 Dreiding, LigScore2 Dreiding, PLP1, PLP2, Jain, PMF and PMF4 used for scoring various poses of the cofactor (NADPH) with the DXR is shown in Table S2. The pose with highest CDOCKER energy was selected as the best pose and it was observed that cofactor firmly binds to active site residues of the DXR protein through hydrogen bonds and hydrophobic interactions (Fig. 8). The pose with the highest CDOCKER energy was selected to represent the studied DXR ligand inside the binding cavity and employed MMFF to minimize the final structure. The NADPH is bound in an extended conformation, through a strong network of hydrogen

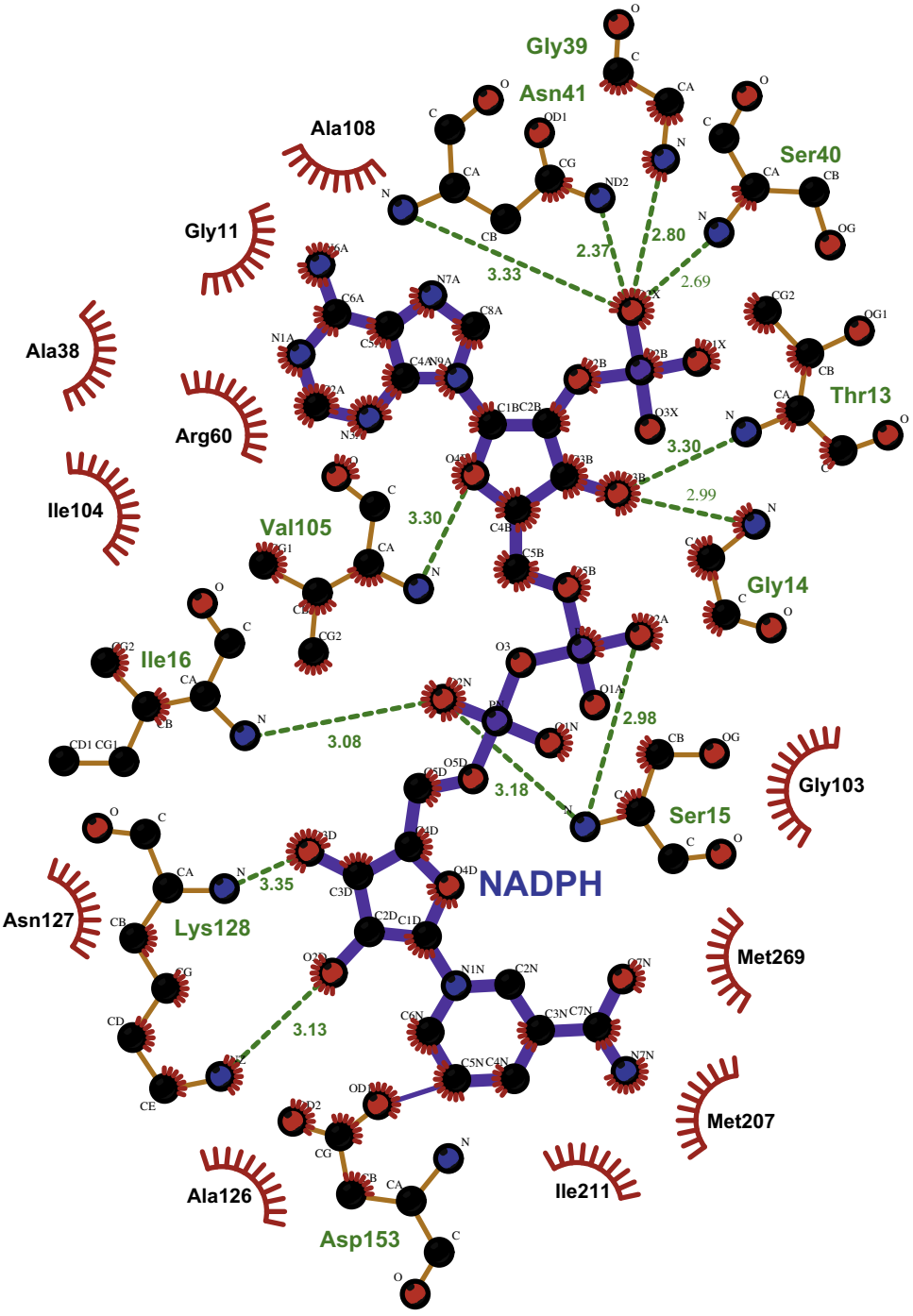

Fig. 8. The docked complex of DXR–NADPH. The figure shows the intermolecular H-bonds and hydrophobic interactions formed between NADPH and DXR amino acids in the NADPH binding pocket that interacted with the cofactor NADPH were shown in 3-D graphical representation. The H-bonds are presented with green dotted lines with atomic distances (Å) with interacting residues. The image was prepared using Ligplus^{+} software.

bonds and hydrophobic contacts (Table 3), as well as a covalent bond via Asp153 residue. The residues involved in DXR–NADPH interactions are listed in Table 3. The H-bond distances were in the range of 2.37–3.35 Å. The phosphate molecule of NADPH forms two H-bonds with Asn41 (α2 helix) with an atomic distance of 3.33 Å and 2.37 Å. Further, it also formed two additional H-bonds with Nitrogen atom of Gly39 and Ser40. Again the N atom of Ser15 (α1 helix) forms two H-bonds with O2A and O2 N atoms of NADPH. Other than these, Thr13, Gly14, Ile16, Val105, Lys128, Asp153 residues forms H-bond with different atoms of NADPH molecule with an average of atomic distance of ~2.37 Å. In addition, a minute observation revealed that when most of the residues which participated in H-bond formation with the cofactor NADPH are discrete amino acids of α-helices. Thus it can conjecture that α-helices plays a vital role in recognition of cofactor with the DXR holo-enzyme. Comparison of the modelled DXR with closest structural homolog showed that Asn41 (NADPH binding residue in template corresponds to Asn41 of ZmDXR is only conserved whereas rest amino acid residues are not conserved indicating different mode of cofactor recognition in citronella. Apart from tight

Table 3
Residues involved in hydrogen bonding and hydrophobic interaction of NDP with DXR obtained using LigPlus+ software.

Properties	No. of hydrogen bonds	Interacting residues forming hydrogen bonds	Hydrophobic interaction forming residues
Modelled DXR	12	Thr13,Gly14, Ser15, Ile16, Gly39, Ser40, Asn41, Val105, Lys128, Asp153	Gly11, Ala38, Arg60, Gly103, Ile104, Ala108, Ala126,Asn127, Met207, Ile211, Met269

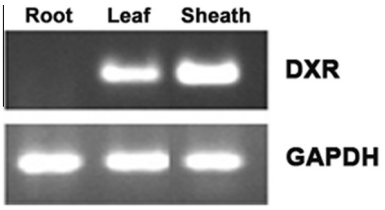

Fig. 9. Semi-quantitative RT-PCR analysis DXR gene in four tissues of citronella plant.

H-bond network the NADPH molecule is tightly anchored in the active pocket of DXR with widespread hydrophobic contacts where most of the residues are from various β-sheets. Moreover, theoretical calculation showed that free energy of binding score (negative CDOCKER energy) of NADPH to DXR is −1043.27 kCal/mol affirms the affinity of DXR model towards NADPH.

2.8. Tissue specific expression study

A distinct differential expression pattern was observed in the semi quantitative RT-PCR analysis in which DXR expression level was found to be highest in leaf sheath followed by leaf (Fig. 9), establishing the fact that DXR is a chloroplast and other plastids specific enzyme which perfectly corroborate with the earlier findings of Lichtenthaler (1999), Lichtenthaler et al. (1997), Rodriguez-Concepcion and Boront (2002) [17,15,37]. As evidence from Fig. 10, it can be clearly observed that maximum expression was seen leaf sheath followed by leaf. Although, in number of previous studies on expression profiling of *dxr* gene in different plants species it was reported that higher level of expression was found in leaf tissues than other tissues tested [38]. But, it can be suggested that high level of mRNA is not enough to reflect the level of enzyme activity in a particular cell type, as mRNA or protein itself is transportable to other site of action. Our result might be an indication that the expression of *dxr* gene is high in plastids abundant in phloem and other cells of sheath tissue. Similar results have also been obtained in a northern blot and immunolocalization experiment conducted in *Catharanthus roseus* for DXR expression [39]. In *C. roseus* the specific expression of four MEP pathway genes including DXR in internal phloem tissues was observed which was unanticipated as photosynthetic pigments are mostly localised in mesophyll tissues. But, the data obtained suggested that internal phloem parenchyma was the main site of isoprenoid biosynthesis and implies the intercellular translocation of metabolites. Recent results suggest that phytol biosynthesis in leaf might involve translocation of an isoprenoid precursor from the vascular tissue in Arabidopsis [40]. Internal phloem plays a key role in formation of precursors and decoration of monoterpenoids. Thus, several mechanisms of transport must exist to deliver isoprenoid intermediates out of the phloem parenchyma plastids and to other compartments and/or cells where they will be further metabolised or stocked [39].

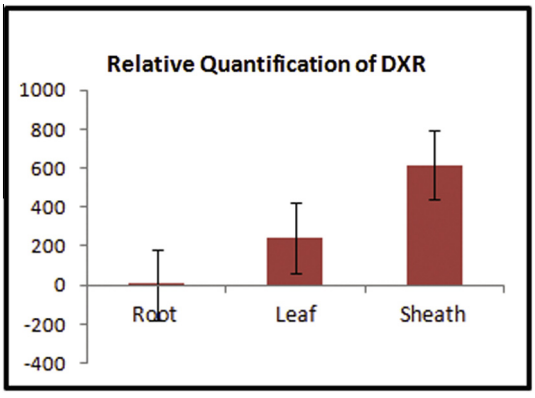

Fig. 10. Comparative expression of DXR gene in three different tissues of mature citronella plant using ΔΔCT method of real time PCR (RQ: Relative Quantification).

3. Conclusion

Citronella is a highly useful medicinal aromatic plant and so increasing its level of secondary metabolite synthesis is of interest to commercial producers of its essential oils. However, the up-regulation of the key enzyme of MEP pathway i.e., DXR will be a significant step for increased essential oil production. This study is the first ever report on characterization of full length DXR in this important medicinal and aromatic plant *C. winterianus*. This study is purely based on *in silico* methods, and so combinatorial methods involving biochemical, enzyme kinetics and site directed mutagenesis will be needed to confirm these predictions. However this is a promising first step towards opening up better avenues for metabolic engineering of secondary metabolite pathways.

4. Materials and methods

4.1. Primary sequence analysis

The primary protein sequence (GenBank Accession number: KJ749651) was subjected to various tools *viz*., InterProScan, SMART (Simple Modular Architecture Research Tool) (http://smart.embl-heidelberg.de), CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) and Pfam database (http://pfam.sanger.ac.uk/) to deduce the protein family DXR gene and to explore the domain arrangement within the protein. In addition, the Protparam tool of ExPaSy proteomic server to analyse the primary structure of the amino acid sequence of DXR protein. Consensus meta-server CONCORD (http://helios.princeton.edu/CONCORD/) was used to assign secondary structure elements from its primary amino acid sequence.

4.2. Phylogenetic study

Primary amino acid sequence of DXR was subjected to BLAST search against non-redundant (NR) database of NCBI. Sequence producing significant alignment (cut-off identity of ⩾85% and *E*-value of 0) with the DXR of citronella were aligned using Multalign and displayed using ESPript. Finally, 2-dimensional phylogenetic tree was established implementing Dayhoff model of substitution using Neighbor-Joining method in MEGA v6.1.0 with a bootstrap value of 1000 iterations.

4.3. Theoretical modelling of DXR

Homology model of DXR protein was generated using MODELLER v9.12 based on suitable template 1R0K-A which was chosen by using fold-recognition server's *viz.,* 3D Jury, Pcons.net, GeneSilico and Geno3D. Further to ensure the correctness in identifying templates for homology modelling through fold recognition servers, DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) was used against protein data bank (PDB). Based on the optimised target-template alignment, modeller facilitated in the development 200 models, of which, the model with the lowest discrete optimized protein energy (DOPE) score and low RMSD (after aligning them on corresponding Cα atom pairs of the template) was selected for further refinements (i.e., loop and side chain refinement) using Discovery Studio3.5 (Accelrys, Inc., San Diego, USA). The refined model was validated by Procheck tool embedded in SAVES server to quantifies the amino acid residues in the available zones of Ramachandran plot to assess the stereochemical quality of the model. Energy profile of modelled DXR was analysed using ProSA tool. The root mean square deviation (RMSD) of the corresponding Cα-atom pairs of both the model and template was calculated using iPBA web server.

4.4. Molecular dynamics simulations

For further refinement the model, modelled DXR was subjected to molecular dynamics (MD) simulations using GROMOS96 43A1 force field in GROMACS 4.6 package [41–43]. For cross validation, we have used two systems i.e., modelled DXR as well its closest structural homolog ZmDXR (template: PDB ID: 1R0K). The systems were solvated with SPC3 water molecules and neutralized by adding equal number of counter ions (Na^+/Cl^-) with a concentration of 0.15 mol. Both the models were placed in an octahedron box maintaining a distance of 1.2 nm from all the directions of protein and periodic boundary. Then the solvated systems were subjected to energy minimization using steepest descent algorithm (2000 steps max.) followed by a 100 ps position restrained MD simulation at constant pressure (1 atm) and temperature (300 K). The long-range electrostatic interactions were computed using the particle-mesh Ewald (PME) [44] with a cut-off value of 0.9 nm for van der Waals and 1.4 nm for electrostatic interactions. All bond lengths were constrained using LINCS algorithm. The trajectories were analysed using GROMACS analysis tools and the structures were analysed using DS3.5 and PyMOL (www.pymol.org).

4.5. Molecular docking of NADPH using CDOCKER

The average coordinate of DXR protein was extracted from stable portion of the trajectory and cofactor NADPH were prepared using CDOCKER module following prepare protocol in DS3.5 prior docking. For docking we adopted the methodology from an earlier study by Dehury et al. (2014) [45]. The ligand was docked within the active site of enzymes and receptors by implementing CHARMm-based molecular dynamics method [46,47] which (binding pocket) was obtained by pair-wise alignment of DXR with the template. Based on the consensus scoring system of CDOCKER module, best ten poses were selected and binding energy of the best 10 poses of NADPH was obtained using "Binding energy module" of DS3.5. The pose with lowest energy, and more number of interacting H-bonds was selected for displaying the enzyme-cofactor interaction using LigPlot$^+$ software.

4.6. Tissue specific expression studies

The differential expression of the DXR gene in three different tissues of citronella (leaf sheath, leaf and root) was tested by semi-quantitative RT PCR and quantitative real time PCR. The primers were designed by primer3 (http://primer3.ut.ee/) with GC% of 60–65% and melting temperature ranging from 60–65 °C. The first strand cDNA was synthesized by using PrimeScript™ 1st strand cDNA synthesis kit (Takara, Clontech) as per manufacturer's instruction. The 10 μl reaction mixtures with 125 ng first strand cDNA with 1X Taq polymerase buffer (Genei, India), 1U Taq polymerase buffer (Genei, India) 10 pmol both forward and reverse primers (Sigma) and 5 mmol dNTP mixure (Invitrogen) were subjected to semi-quantitative RT PCR in GeneAmp Thermo Cycler (Applied Biosystem, USA). The amplification condition were 5 min at 94 °C, followed by 35 cycles of 1 min at 94 °C 1 min at 67.5 °C 1 min at 72 °C and final extension was done for 10 min at 72 °C. The amplified products were visualized in 2% agarose gel stained with EtBr. In this study rice GAPDH gene was used as reference. The relative expression of DXR gene in the same 3 tissues were checked with quantitative real time PCR by ΔΔCT method [48] on StepOnePlus Real-Time PCR System (Applied Biosystem) using the SuperScript III Platinum SYBR Green One Step qRT-PCR with ROX Kit (Invitrogen, CA, USA) according to manufacturer's instruction. PCR amplification was performed under the following conditions: 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min finally melting at 95 °C for 15, 60 °C for 1 min. The gene expressions were normalized against an internal reference gene, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and root tissue was arbitrarily chosen to be the calibrator of tissue gene expression.

Acknowledgements

K.D. and P.S. conceived and designed the project, K.D., B.D., M.P. and M.K.M. acquired the data, K.D., B.D., M.P. and P.S. analysed and interpreted the data, K.D., B.D. and P.S. wrote the paper. We are thankful to Ms. Lochna Patar and Ms. Mausumi Sahu for their valuable help in technical aspects of this work. We are highly grateful to Dr. Iswar Boruah for providing plant materials for the experiment. We would like to thank the Distributed Information Center, Dept. of Agricultural Biotechnology, AAU, Jorhat for providing computational facility. We deeply acknowledge the Department of Science and Technology, Government of India for Inspire Fellowship (IF120251) as well as for financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fob.2015.04.005.

References

[1] Shasany et al. (2000) Phenotypic and RAPD diversity among *Cymbopogon winterianus* Jowitt accessions in relation to *Cymbopogon nardus* Rendle. Genet Resour Crop Evol 47 (5), 553–559.
[2] Silva, C.F., Moura, F.C., Mendes, M.F. and Pessoa, F.L.P. (2011) Extraction of citronella (*Cymbopogon nardus*) essential oil using supercritical CO_2: experimental data and mathematical modelling. Braz J Chem Eng 28 (2), 343–350.
[3] Thorsell, W., Mikiver, A. and Tunon, H. (2006) Repelling properties of some plant materials on the tick *Ixodes ricinus* Lin. Phytomedicine 13, 132–134.
[4] Katz, T.M., Miller, J.H. and Hebert, A.A. (2008) Insect repellents: historical perspectives and new developments. J Am Acad Dermatol 58 (5), 865–871.
[5] Simic et al. (2008) Essential oil composition of *Cymbopogon winterianus* and *Carum carvi* and their antimicrobial activities. Pharm Biol 46 (6), 437–441.
[6] deBillerbeck et al. (2001) Effects of *Cymbopogonnardus* (L.) W. Watson essential oil on the growth and morphogenesis of *Aspergillusniger*. Can J Microbiol 47 (1), 9–17.
[7] Nakahara et al. (2003) Chemical composition and antifungal activity of essential oil from *Cymbopogon nardus* (citronella grass). Jpn Agric Res Q 37 (4), 249–252.
[8] www.webmd.com/vitamins-supplements/ingredientmono-627-citronella%20oil.aspx, Retrieved on: 17th August, 2014.

[9] Cassel, E. and Vargas, R.M.F. (2006) Experiments and modeling of the *Cymbopogon winterianus* essential oil extraction by Steam Distillation. J MexChemSoc 50 (3), 126–129.
[10] Rohmer et al. (1993) Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. Biochem J 295, 517–524.
[11] Sprenger et al. (1997) Identification of a thiamine-dependent synthase in *Escherichia coli* required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamine and pyridoxol. Proc Natl Acad Sci U S A 94, 12857–12862.
[12] Eisenreich, W., Bacher, A., Arigoni, D. and Rohdich, F. (2004) Biosynthesis of isoprenoids via the non-mevalonate pathway. Cell Mol Life Sci 61, 1401–1426.
[13] Bouvier, F., Rahier, A. and Camara, B. (2005) Biogenesis, molecular regulation and function of plant isoprenoids. Prog Lipid Res 44, 357–429.
[14] Lichtenthaler, H.K., Rohmer, M. and Schwender, J. (1997) Two independent bio-chemical pathways for isopentenyl diphosphate (IPP) and isoprenoid biosynthesis in higher plants. Physiol Plant 101, 643–652.
[15] Lichtenthaler, H.K., Schwender, J., Disch, A. and Rohmer, M. (1997) Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate independent pathway. FEBS Lett 400, 271–274.
[16] Zeidler, J.G., Lichtenthaler, H.K., May, H.U. and Lichtenthaler, F.W. (1997) Is isoprene emitted by plants synthesized via the novel isopentenylpyrophosphate pathway? Z Naturforsch 52, 15–23.
[17] Lichtenthaler, H.K. (1999) The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol 50, 47–65.
[18] Rohmer, M. (1999) Themevalonate-independent methylerythritol 4-phosphate (MEP) pathway for isopreniod biosynthesis, including carotenoids. Pure Appl Chem 71, 2279–2284.
[19] Sangari et al. (2010) A new family of enzymes catalyzing the first committed step of the methylerythritol 4-phosphate (MEP) pathway for isoprenoid biosynthesis in bacteria. Proc Natl Acad Sci U S A 107 (32), 14081–14086.
[20] Cordoba, E., Salmi, M. and Leo, P.N. (2009) Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. J Exp Bot 60 (10), 2933–2943.
[21] Kuzuyama, T. (2002) Mevalonate and non-mevalonate pathways for the biosynthesis of isoprene units. Biosci Biotechnol Biochem 66, 1619–1627.
[22] Kuzuyama, T., Shimizu, T., Takahashi, S. and Seto, H. (1998) Fosmidomycin, a specific inhibitor of 1-deoxy-D-xylulose 5-phosphate reductoisomerase in the nonmevalonate pathway for terpenoid biosynthesis. Tetrahedron Lett 39, 7913–7916.
[23] Takahashi, S., Kuzuyama, T., Watanabe, H. and Seto, H. (1998) A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. Proc Natl Acad Sci U S A 95, 9879–9884.
[24] Lorenzo, D. et al. (2000) Composition and stereo analysis of *Cymbopogon winterianus* Jowitt oil from Southern Brazil. Flavour Fragr J 15, 177–181.
[25] Xing et al. (2010) Disruption of the 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) gene results in albino, dwarf and defects in trichome initiation and stomata closure in Arabidopsis. Cell Res 20 (6), 688–700.
[26] Grolle, S., Bringer-Meyer, S. and Sahm, H. (2000) Isolation of the *dxr* gene of *Zymomonas mobilis* and characterization of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase. FEMS Microbiol Lett 191, 131–137.
[27] Gong et al. (2005) Molecular cloning and characterization of a 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene from *Ginkgo biloba*. DNA Seq 16 (2), 111–120.
[28] Khemvong, S. and Suvachittanont, W. (2005) Molecular cloning and expression of a cDNA encoding 1-deoxy-D-xylulose-5-phosphate synthase from oil palm *Elaeis guineensis* Jcaq. Plant Sci 169, 571–578.
[29] Yao et al. (2008) Molecular cloning, expression profiling and functional analysis of a *DXR* gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase from *Camptotheca acuminate*. J Plant Physiol 165, 203–213.
[30] Yan et al. (2009) Molecular characterization and expression of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) gene from *Salvia miltiorrhiza*. Acta Physiol Plant 31, 1015–1022.
[31] Carretero-Paulet et al. (2002) Expression and molecular analysis of the Arabidopsis DXR gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. Plant Physiol 129 (4), 1581–1591.
[32] Lange et al. (2011) Improving peppermint essential oil yield and composition by metabolic engineering. Proc Natl Acad Sci U S A 108 (41), 16944–16949.
[33] Mahmoud, S.S. and Croteau, R.B. (2001) Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. Proc Natl Acad Sci U S A 98 (15), 8915–8920.
[34] Emanuelsson, O., Nielsen, H. and von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8, 978–984.
[35] vonHeijne, G., Stepphuhn, J. and Herrmann, R.G. (1989) Domain structure of mitochondrial and chloroplast targeting peptides. Eur J Biochem 180, 535–545.
[36] Ricagnoa et al. (2004) Crystal structure of 1-deoxy-D-xylulose-5-phosphate reductoisomerase from *Zymomonas mobilis* at 1.9-Å resolution. Biochim Biophys Acta 1698, 37–44.
[37] Rodriguez-Concepcion, M. and Boront, A. (2002) Elucidation of the methylerithritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieve through genomics. Plant Physiol 130, 1079–1089.
[38] Sangwan et al. (2013) De novo sequencing and assembly of *Centella asiatica* leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism. Gene 525, 58–76.
[39] Burlat et al. (2004) Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpeneindole alkaloids and isoprenoid-derived primary metabolites. Plant J 38, 131–141.
[40] Nagata, N., Suzuki, M., Yoshida, S. and Muranaka, T. (2002) Mevalonic acid partially restores chloroplast and etioplast development in Arabidopsis lacking the non-mevalonate pathway. Planta 216, 345–350.
[41] Berendsen, H.J.C., Van Der Spoel, D. and Van Drunen, R. (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 91, 43–56.
[42] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., et al. (2005) GROMACS: fast, flexible, and free. J Comput Chem 26, 1701–1718.
[43] Dehury et al. (2014) Structure-based computational study of two disease resistance gene homologues (Hm1 and Hm2) in maize (*Zea mays* L.) with implications in plant-pathogen interactions. Plos One 9, e97852.
[44] Lindahl, E., Hess, B. and Van Der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7, 306–317.
[45] Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: an N. log(N) method for Ewald sums in large systems. J Chem Phys 98, 10089–10092.
[46] Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., et al. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4, 187–217.
[47] Guosheng et al. (2003) Detailed analysis of grid-based molecular docking: a case study of CDOCKER—a CHARMm-Based MD docking algorithm. J Comput Chem 24, 1549–1562.
[48] Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta DeltaC(T)) method. Methods 25 (4), 402–408.